

AD \_\_\_\_\_  
(Leave blank)

Award Number: W81XWH-12-1-0444

TITLE: Genomic Characterization of Brain Metastasis in Non-Small Cell Lung Cancer Patients

PRINCIPAL INVESTIGATOR: Mark A. Watson, MD PhD

CONTRACTING ORGANIZATION:  
Washington University  
Saint Louis, MO 63130-4862

REPORT DATE: 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

X Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE January 2014		2. REPORT TYPE Final		3. DATES COVERED 1 September 2012 - 31 October 2013	
4. TITLE AND SUBTITLE Genomic Characterization of Brain Metastasis in Non-Small Cell Lung Cancer Patients				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-12-1-0444	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Mark A. Watson, MD PhD  E-Mail: researchgrants@wusm.wustl.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Washington University Saint Louis, MO 63130-4862				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Despite advances in systemic therapy, brain metastases remain a significant cause of mortality in non-small cell lung cancer (NSCLC) patients. We hypothesize that subpopulations of primary NSCLC tumor cells evolve through a multistep process of genomic and epigenomic alterations that result in a metastatic cell phenotype. In this proposal, we have utilized 'next generation' exome sequencing to perform a comparative analysis of the genomes of patient-matched primary NSCLC and brain metastatic tumor cell populations. A population of 12 patients with NSCLC were used for a discovery set. We identified genomic alterations that were enriched in metastatic tumor cell populations and recurrent across patients. Several of these alterations (PIK3CA E535K, MAPK4 P246T) have been previously documented in NSCLC and are potential mediators of targeted therapeutics. One alteration (FES E651G) was correlated with time to brain metastatic recurrence in an independent set of primary NSCLC patients. This pilot data set demonstrate the validity and potential clinical utility of this experimental approach. Newly funded efforts are expanding these studies to larger patient populations, with the goal of identifying set of genomic alterations that define a gene network-based predictor of the brain metastatic phenotype in early stage NSCLC patients.					
15. SUBJECT TERMS- Non-small cell lung cancer; brain metastasis; genomics					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	10	19b. TELEPHONE NUMBER (include area code)

## Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	5-7
Key Research Accomplishments.....	8
Reportable Outcomes.....	9
Conclusions.....	10

## **INTRODUCTION**

Despite advances in systemic therapy, brain metastasis remains a significant cause of mortality in non-small cell lung cancer (NSCLC) patients. Nearly 50% of patients with NSCLC will develop brain metastases during the course of their disease. If patients at high risk for brain metastasis could be selectively identified, then the benefits of more frequent MRI-based screening, PCI, or other systemic therapies could outweigh the cost and morbidities associated with aggressive screening and therapy in a population of otherwise early stage patients undergoing 'curative' surgical resection. Moreover, if subclonal populations of potential metastatic cells harbor unique and identifiable molecular alterations, specific targeted therapies could be studied to prevent or delay metastasis to the brain. In this pilot project, we have demonstrated how a comparative genomic analysis of primary NSCLC tumors and their brain metastatic derivatives reveal complex but possibly recurrent patterns of genomic alteration in early stage primary tumors that may be predictive of eventual brain metastasis. Ultimately, these findings may reveal opportunities for targeted therapeutics that are designed against specific subpopulations of tumor cells that demonstrate metastatic potential.

## BODY (PROJECT SUMMARY)

Below, we summarize the work performed during this one-year project, based upon the stated tasks outlined in our proposal (with particular emphasis on Task 3). Because of the plummeting costs of DNA sequencing and the efficiency with which we were able to complete Tasks 1-3, we were able to perform additional studies (described below) that were performed as an extension of Task3d.

### Task 1. Prepare samples for Next Generation Sequencing (months 1-2)

We identified over 24 institutional cases of NSCLC patients from whom both primary lung tumor tissue and brain metastatic tumor tissue were physically available. We applied the following selection criteria to 'qualify' cases: 1) Tumor tissue block from both primary tumor and brain metastatic lesion contained sufficient tissue; 2) Tumor tissue contained at least 50% tumor nuclei cellularity; 3) Tumor tissue contained less than 20% necrosis. Based on these criteria, 12 cases (24 tissues) were sectioned and used to isolate genomic DNA (**Table 1**). This task was the biggest challenge of the project and continues to limit follow up investigations. Namely, it is difficult to identify paired tumor tissue specimens of primary NSCLC and paired brain-metastatic tissue of sufficient tissue adequacy and tumor cellularity.

Patient	Histology	Time to brain met
1	AD	29 mo
2	SCC	7 mo
3	AD	Synch
4	SCC	6 mo
5	ASQ	7 mo
6	SCC	17 mo
7	SCC	6 mo
8	AD	Synch
9	AD	7 mo
10	AD	Synch
11	AD	18 mo
12	AD	Synch

**Table 1. Paired primary / brain metastasis cases used for preliminary studies**

### Task 2. Conduct Next Generation Sequencing on 24 DNA samples (months 3-4)

DNA from paired primary tumor and brain metastatic tissue was prepared. One microgram of each DNA was fragmented and used for NGS library preparation. For this pilot project, we did not utilize matching non-malignant, constitutional DNA from each case as a reference. We believed that this would be cost prohibitive and would be unnecessary, as we were primarily interested in identifying alterations that were enriched in metastatic tumor vs. primary tumor, which by definition would imply that identified mutations were somatic anyway. We also believed that constitutional ('germline') SNPs could be eliminated from

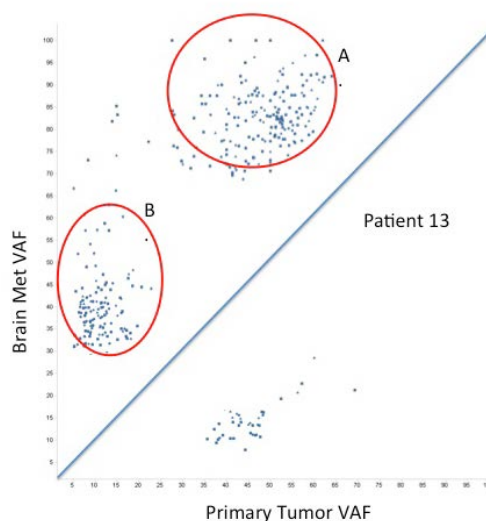
Patient	Tissue Source	Platform	Total Reads	Mapped reads	%	% Coverage > 25X	cVariants
PT_2	Primary	V4	133,766,317	37,616,520	76	0.75	39,416
	Metastasis		125,692,786	40,670,007	89	0.77	38,691
PT_3	Primary	V4	133,160,791	34,383,270	89	0.67	36,826
	Metastasis		144,063,947	43,595,551	89	0.79	39,308
PT_6	Primary	V4	132,585,500	42,616,365	91	0.70	35,577
	Metastasis		135,869,579	37,537,776	91	0.68	34,686
PT_9	Primary	V3	152,657,352	79,033,802	75	0.87	38,744
	Metastasis		100,815,248	58,692,374	81	0.84	37,025
PT_10	Primary	V3	82,912,528	48,894,061	83	0.78	36,479
	Metastasis		86,340,606	55,417,833	82	0.79	38,642
PT_11	Primary	V3	100,254,966	50,192,271	77	0.85	39,941
	Metastasis		99,605,064	56,830,286	79	0.88	40,360
PT_12	Primary	V3	100,323,762	45,729,694	76	0.84	40,244
	Metastasis		95,041,794	57,053,178	81	0.86	38,592
PT_13	Primary	V4	129,394,368	41,221,078	90	0.78	46,877
	Metastasis		138,426,598	45,623,392	89	0.82	46,178

**Table 2. Exome sequencing statistics.** Representative data for 8 of the 12 sample pairs are shown.

consideration by comparing with 1,000 Genomes and other reference databases. As we learned, it is still of great benefit to include reference constitutional DNA sequence data in the analysis of each case, and will do this going forward, particularly as sequencing prices have dropped considerably. Exome libraries were prepared using Agilent SureSelect V3 or V4 exome capture kits and sequenced on an Illumina HiSeq sequencer. Sequencing statistics are presented in **Table 2**. Approximately 78% of all reads had a coverage depth of greater than 25X. Given the level of sample multiplexing performed and the fact that all DNA was extracted from formalin fixed, paraffin embedded tumor samples, some more than 10 years old, the overall sequence quality was very high.

### Task 3. Analyze Next Generation Sequencing Data (months 5-12)

Using paired exome capture sequencing of 9 patients from our pilot set, we identified genomic variants (SNVs and CNVs to date) that are enriched in metastatic tumors. Aligned reads were used for variant calling using the VARSCAN tool, setting parameters to require at least 5 high-quality reads for variant support and coverage of at least 25X over the called variant base. During alignment all reads mapping to multiple genome locations were excluded from the analysis to prevent false variant calls. Variants represented in the 1,000 Genomes database, Exome Variant Server (EVS, <http://evs.gs.washington.edu>), or a local database of known platform-specific variant artifacts were excluded from further analysis. From this analysis, we identified a total of 7,447 variants among all 9 sample pairs and a total of 416 individual SNVs from high quality reads that were represented in more than one patient and were not present in any reference genome database. For each sample pair, we specifically looked for variants that either were uniquely called in the metastatic tumor relative to the patient-matched primary tumor or observed at enhance variant allele frequency (VAF). **Figure 1** demonstrates a VAF plot for one patient, illustrating variants that are detected at low VAF frequency in the primary tumor but greatly enriched in the metastatic lesion, or that are present in the heterozygous state in primary tumors and demonstrate subsequent loss of heterozygosity in the metastasis. Interestingly, the number of metastasis-enhanced (ME) variants identified varied greatly from patient to patient (range 1-87), although this number did not appear to immediately correlate with either sequencing quality metrics or clinical parameters, such as time to metastasis. Across all 9 cases, we found a total of 144 somatic gene mutations with enriched allele frequency in metastasis vs. primary tumor. Although none of these mutations or genes was recurrent in more than one of the 9 samples, several of them, including *PIK3CA* E545K (not detectable at 25X read depth in patient 1 primary tumor, but present at 52% VAF in the corresponding metastasis) and *MAPK4* P246T (present at 6% VAF in patient 9 primary tumor and enriched to 63% in the corresponding metastasis), have been previously identified in NSCLC tumor genomes and are potential modulators of targeted therapeutics. We believe that similar analyses of larger, more uniformly defined case sets coupled with gene network analyses, will identify specific sets of genes whose somatic mutation will serve as clonal markers of metastatic progression.

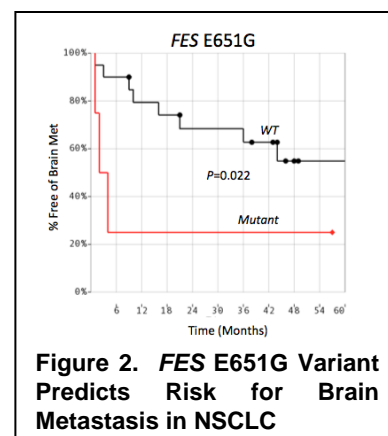


**Figure 1. Variant Allele Frequency (VAF) Plot of Patient 13 primary tumor vs. metastasis.** Populations of variants that are enriched in the brain metastasis vs. the primary tumor (B) or that demonstrate loss of heterozygosity, presumably loss of a wild type allele (A) are indicated. Variants present at equivalent VAF in both samples have been filtered.

Gene	AA Change	COSMIC	GERP	PolyPhen	Validation			
					Frequency in Primary Exome (n=9)	Frequency in No Met (n=11)	Frequency in Brain Met (n=13)	Frequency in Other Met (n=17)
FKBP9	H567Q		5.07	D	55%	36%	31%	47%
FES	E651G		5.42	P	44%	9%	23%	18%
FOXD4L1	I155V	Y	2.57	D	33%	0%	23%	23%
CCDC37	E273*		4.59	NA	22%	0%	23%	0%
PDLIM2	T597P		4.72	B	22%	9%	23%	0%
BAGE2	D40N		NA	NA	22%	0%	15%	18%
KRAS	G12V	Y	5.68	D	22%	9%	8%	6%

**Table 3. Validated, recurring variants in primary NSCLC with and without eventual brain metastasis.** Gene and specific amino acid change caused as a result of the variant are shown, along with GERP and PolyPhen tool predictions. Frequency of the variant in the 9 cases subjected to exome sequencing is shown as well as frequencies in the validation patient cohorts without or with eventual metastasis specifically to brain or other distant organ site.

Since none of the ME variants identified in our pilot study were recurrent in more than one of the nine samples analyzed, we also looked for variants that were present in at least the primary tumor specimen in more than one case of this phenotypically ‘extreme’ and homogeneous cohort. Based on the variant filtering steps described above, we were surprised to find 416 specific variants that were recurrent in more than one primary NSCLC from this 9 patient cohort. While we expected that many of these were artifacts, despite several filtering strategies and manual review of mapped sequence reads, we were able to select 48 ‘high confidence’ variants for validation in an independent set of samples, using targeted amplicon sequencing. As shown in **Table 3**, seven specific variants that were identified in more than one of the original 9 cases analyzed by exome sequencing were also detected recurrently in an independent set of 41 primary NSCLC (adenocarcinoma) tumors with or without brain or other organ metastasis. For reference, the canonical *KRAS* G12V mutation was identified in 2 of 9 cases originally analyzed by exome sequencing and approximately 7% of the validation cases, regardless of whether patients experienced a brain relapse or not. Variants such as *CDC37* E273\* were specifically found in those patients who developed brain metastasis, while other variants such as *BAGE2* D40N were detected in patients with both brain and other distant organ site metastasis. Although the number of mutation positive cases was still too low in this cohort to achieve meaningful statistical significance for most other recurrent variants, we did confirm that the *FES* non-receptor tyrosine kinase oncogene, E651G variant correlated with risk of brain metastasis ( $p < 0.02$ , Wilcoxon-Gehan test), even in this small sample set (**Figure 2**). It is remarkable that this same amino acid is mutated in multiple patients, and the location of E651G variant in the kinase domain of the protein together with its highly non-conservative substitution suggest that it is likely to be functionally significant.



## **KEY RESEARCH ACCOMPLISHMENTS**

- Identified 24 cases of paired primary NSCLC / brain metastasis for sequence analysis.
- Performed exome sequencing on 12 cases of paired primary NSCLC / brain metastasis.
- Identified candidate mutations that were enriched in brain metastasis relative to primary tumor.
- Identified candidate mutations that were recurrent in the primary tumors of NSCLC patients who developed brain metastasis.
- Validated that the presence of at least 1 mutation (*FES* E651G) in primary NSCLC correlates with time to metastasis in a small validation cohort.



## **REPORTABLE OUTCOMES**

- Used this preliminary data to successfully compete for additional NIH / NCI R01 funding (1R01CA182746- 4th percentile; Impact Score 19; Anticipated funding 7/2014).

## **CONCLUSIONS**

Although many of the biological pathways, processes, and key genetic components associated with solid tumor metastasis have been well defined, these advances have not yet led to robust clinical biomarkers for predicting metastatic behavior in primary NSCLC. Most prognostic markers evaluated to date are based on gene expression or immunohistochemical staining of the primary tumors themselves, but alterations of single or a few genes do not seem to reliably predict brain metastasis in patients with resected early stage NSCLC. This is perhaps not surprising given the complex genomic alterations seen typically in NSCLC and the overall complexity of the metastatic process. Furthermore, if only a small fraction of malignant cells in the primary tumor harbor genomic or transcriptional changes that predispose them to metastasize, it may not be possible to directly identify these from primary tumor analyses alone. We believe that a more powerful approach will be to directly compare the genomes and/or transcriptomes of primary tumors and subsequent metastases from the same patient, in order to directly identify molecular pathways that are specific to or enriched in metastatic cell populations and that could provide insight into therapeutic sensitivity and resistance in recurrent, metastatic disease. The study summarized in this pilot project demonstrates, with a small number of patients, the feasibility of the technical and analytical approach. With only 9 cases analyzed, we have identified several candidate biomarker gene mutations that have passed at least one round of validation in an independent cohort of patients. More importantly, follow-up work that will be performed over the next several years using larger number of patients, combined gene expression and gene mutational profiling, and a computational approach to gene-networked based biomarker discovery, should provide more robust genomic biomarkers that can predict this specific phenotype (brain metastasis) and offer possibly novel therapeutic targets to patients with early stage NSCLC.